R. Prentice

Surrogate Endpoints in Chronic Disease Prevention: How Comprehensive Do They Need to Be?

Ross Prentice
Fred Hutchinson Cancer Research Center

- Outcome selection for the evaluation of chronic disease prevention interventions
- Surrogate outcome criteria (and validation)
- Examples of breast screening and hormone replacement therapy interventions

Outcome Selection in CT's to Evaluate Chronic Disease Prevention Interventions

- Ultimately wish to answer whether or not intervention be recommended to pertinent subsets of the general population for disease prevention/health promotion purposes?
- Inherent focus on overall benefits versus risks in randomized, controlled trials
- Primary and secondary outcomes and potential adverse outcomes typically designated in an attempt to 'capture' the clinical outcomes that are most plausibly affected (beneficially or adversely) by the intervention

  e.g. Women's Health Initiative Clinical Trial

  - Low fat eating pattern: breast and colorectal cancer; CHD
  - Hormone replacement therapy: CHD; other CVD; fractures; breast cancer
  - Calcium and Vitamin D: hip fractures; other fractures; colorectal cancer

- Primary, secondary and potential adverse outcomes themselves surrogate for summary measure of risk versus benefit (total mortality; quality of life; . . .)
- Most discussion of surrogate outcomes have focused on their ability to yield valid information about intervention effects on the primary outcome, but surrogate outcome trials typically yield little information about other relevant clinical outcomes.

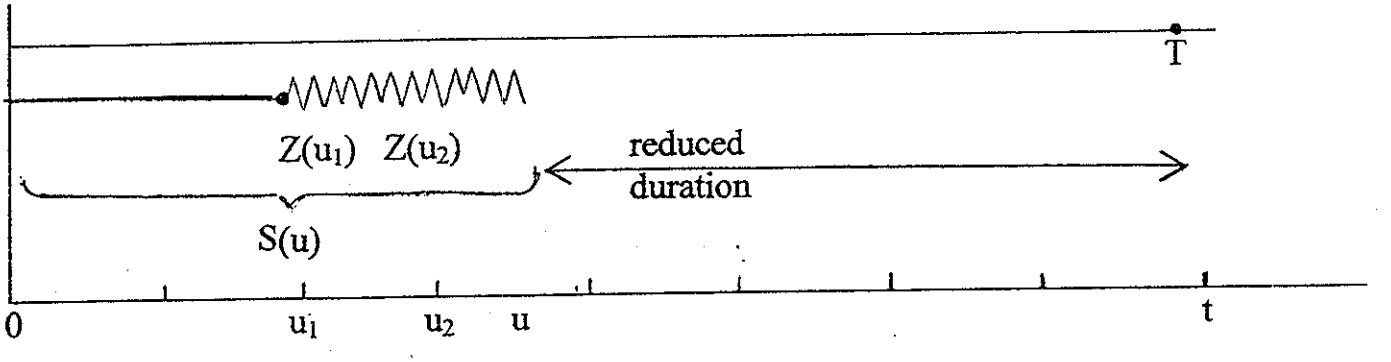
Surrogate Outcome Definition and Criteria

Prentice, 1989, *Statist. in Med.*

T - time from randomization to true (primary) outcome

$S(t) = \{Z(u); 0 \leq u < t\}$ history up to time t of possibly vector-valued surrogate process

x - intervention (treatment) indicator variable (e.g. $x = 0$ for control; $x = 1$ for intervention)



Definition: S a surrogate for T for evaluating intervention x if $\lambda_T(t; x) \equiv \lambda_T(t) \Leftrightarrow \text{pr}\{S(t); x, F(t)\} \equiv \text{pr}\{S(t); F(t)\}$

Surrogate Outcome Definition

$$\lambda_T(t;x) \equiv \lambda_T(t) \;\Leftrightarrow\; \text{pr}\{S(t);x,F(t)\} = \text{pr}\{S(t);F(t)\}$$

Surrogate Outcome Criteria

$$\begin{aligned}\lambda_T(t;x) &= E\{\lambda_T(t);S(t),x\} \\ &= \int \lambda_T\{t;S(t),x\}\ \text{pr}\{S(t);x,F(t)\}\end{aligned}$$

* (i) $\lambda_T\{t;S(t),x\} \equiv \lambda_T\{t;S(t)\}$

(ii) $\lambda_T\{t;S(t)\} \not\equiv \lambda_T(t)$

(iii) $E[\lambda_T\{t;S(t)\} \mid x,F(t)] \not\equiv E[\lambda_T\{t;S(t) \mid F(t)\}]$

(iv) Missing data rates for S given x are independent of T

Surrogate Outcome Criteria

(i) $\lambda_T\{t; S(t), x\} \equiv \lambda_T\{t; S(t)\}$

- S fully mediates the relationship between x and T
- Can be empirically examined by assessing the extent to which an intervention or treatment effect of x on T is explained by including the surrogate variable histories in the analysis

  e.g. percentage treatment effect explained (PTE)

  Freedman, Graubard and Schatzkin (1992). *Statist. in Med.* - T binary

  Lin, Fleming and DeGruttola (1997). *Statist. in Med.* - T failure time
- Large sample sizes may be required to rule out moderate departures from (i)
- Biological/mechanistic knowledge important to assessing plausibility of (i), but unlikely to establish (i) in situations where relationship of x to T is uncertain
- May be possible to avoid the need to assess (i) if meta-analyses allow direct study of concordance of tests of the two null hypotheses (e.g. Daniels and Hughes, *Statist. in Med.*, 1997)

Surrogate Outcome Criteria (cont'd)

(ii) $\lambda_T\{t; S(t)\} \not\equiv \lambda_T(t)$

- S is a risk factor for T
- The stronger this relationship, the closer the correspondence between the two null hypotheses

(iii) $\int \lambda_T\{t; S(t)\}\ \mathrm{pr}\{S(t); x, F(t)\} \not\equiv \int \lambda_T\{t; S(t)\}\mathrm{pr}\{S(t); F(t)\}$

- Requires 'orderly' relationship between S and x so that departure from null hypothesis for S has some implication for T
- See Buyse and Molenbergh's (1998, *Biometrics*) for an illustration with binary T and $S\varepsilon\{0, 1, 2\}$

(iv) Missing data rates for S given x are independent of T

- Needed so that representative data are available to test

$$\mathrm{pr}\{S(t); x, F(t)\} \equiv \mathrm{pr}\{S(t); F(t)\}$$

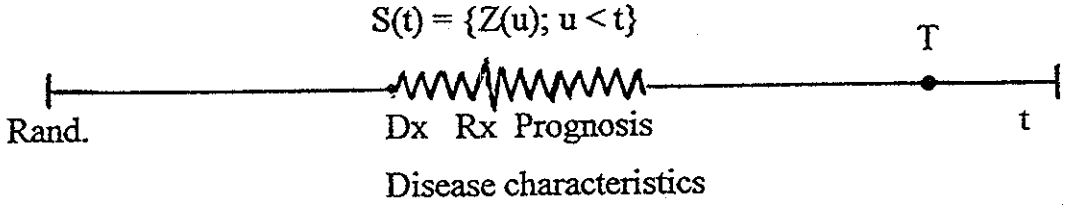
Example 1: x - a specified program of mammographic screening among women ≥ 50

Is health improved/longevity increased?

T - time from randomization to breast cancer mortality

Possible choices for S

- Time from randomization to breast cancer incidence
  - Disaster even though $x \rightarrow S \rightarrow T$
  - Care needed if ascertainment of S affected by x
- Time from randomization to advanced breast cancer diagnosis
  - Condition (i) requires full sensitivity and specificity with regard to breast cancer mortality risk



Note: Instead of replacing T this surrogate process could be regarded as providing auxiliary data for study subjects having censored T values, toward a more powerful test of $\lambda_T(t; x) \equiv \lambda_T(t)$.

Example 2: Hormone replacement therapy (HRT) among postmenopausal women

Do health benefits exceed risks? (Among women with history of heart disease? Among healthy women?)

Does HRT reduce coronary heart disease?

(i) Many observational studies suggesting 40–50% reduction in CHD; and duration dependent 20–30% increase in breast cancer; and recently 2–4 fold increase in various thrombotic disease

(ii) Postmenopausal Estrogen/Progestin Interventions Trial (*JAMA*, 1995)

875 healthy postmenopausal women to assess effects of HRT on heart disease risk factors (HDL-C; LDL-C; SPB; serum insulin; fibrinogen) as well as on endometrial histology and bone mineral density

(iii) Heart and Estrogen/Progestin Replacement Study (HERS) (*JAMA*, 1998)

2763 postmenopausal women with established coronary disease

.625 mg/day CEE plus 2.5 mg/day CEE

- No difference in CHD events over 4.1 year follow-up period RR=.99(.80, 1.22)

Year 1	1.52 (1.01, 2.29)
Year 2	1.00 (0.67, 1.49)
Year 3	0.87 (0.55, 1.37)
Years 4 and 5	0.67 (0.43, 1.04)

Discussion

- Even interventions that may have been widely used for long periods of time may have important unrecognized effects
- Efforts to reach conclusions about intervention effects on clinical endpoints by studying intervention effects on short term markers may be unlikely to achieve their goal
  - Any such efforts may require a large battery of markers and may suffer from an inability to meaningfully analyze a high-dimensional surrogate
- Surrogate outcome concepts (i.e. extent of mediation of intervention effects, PTE) may be quite valuable in helping to guide a research agenda wherein preventive interventions are tested in relation to pertinent outcomes of increasing clinical relevance.
- Auxiliary data analysis concepts may allow some strengthening of intervention comparisons at various steps in such an agenda

- Methdologic work needed to
  - Provide PTE estimates for various (S, T) choices in relation to key interventions or classes of interventions
  - Develop flexible data analysis methods for such estimation (e.g. to accommodate measurement error in S; to efficiently analyze multivariate failure time data)